

Graph spectra as a systematic tool in computational biology

Anirban Banerjee, Jürgen Jost*

Max Planck Institute for Mathematics in the Sciences,
Inselstr.22, 04103 Leipzig, Germany,

banerjee@mis.mpg.de, jost@mis.mpg.de

February 5, 2008

Abstract

We present the spectrum of the (normalized) graph Laplacian as a systematic tool for the investigation of networks, and we describe basic properties of eigenvalues and eigenfunctions. Processes of graph formation like motif joining or duplication leave characteristic traces in the spectrum. This can suggest hypotheses about the evolution of a graph representing biological data. To this data, we analyze several biological networks in terms of rough qualitative data of their spectra.

1 Introduction

Modern biological data are often represented in terms of graphs. Microarray data may lead to graphs whose vertices are genes and whose edges stand for correlations, hypothetically interpreted as interactions. In the study of the proteome of a cell, one sees protein-protein interaction networks. Likewise, at a higher level, cell-cell interactions naturally lead to interaction graphs. A particular example are neural networks where the vertices stand for neurons and the edges for synaptic connections. In populations, graphs encode networks of interactions between individuals, and in ecosystems, trophic and other interactions between species. A special case are phylogenetic trees that express descendance relations between species.

The natural question then is how biological content can be extracted from these formal structures, the graphs to which the biological data are reduced. In graph theory, many concepts have been developed that capture various quantitative

*This author thanks the local organizers Bülent Karasözen and Gerhard Wilhelm Weber for an intellectually stimulating meeting and outstanding hospitality.

or qualitative aspects of a graph (for an algebraic, graph theoretical approach, see e.g. [14, 9], for statistical mechanics methods see e.g. [1, 24, 11]). Recently, a power law behavior of the degrees has become quite popular as it seems to be rather ubiquitous in biological and other data ([7]). Another powerful invariant of the graph is its first eigenvalue that provides estimates for how difficult it is to cut up the graph into disjoint components (see [10], or for how easily dynamics at the vertices can get synchronized ([26, 27, 20, 4]) and many other articles). Useful as any such individual invariant may be, however, it cannot capture all the qualitative aspects of a graph. For example, graphs with the same degree distribution can have a completely different synchronizability ([2, 3]). Also, by their very nature, universal properties like a power law degree distribution capture what is common to large classes of graphs, but fail to identify what is specific about graphs from a particular domain, and what distinguishes those graphs qualitatively from those from other fields.

Therefore, in this contribution, we advocate a set of graph invariants that, on one hand, can be easily graphically represented and therefore visually analysed and compared, and on the other hand, yields an essentially complete qualitative characterization of a graph. This is the spectrum of the graph Laplacian ([21, 23, 18, 19, 33, 8, 6]). In [5], we have applied this method to the study of protein-protein interaction networks.

2 The spectrum of a graph

Let Γ be a finite and connected graph with N vertices. Vertices $i, j \in \Gamma$ that are connected by an edge of Γ are called neighbors, $i \sim j$. The number of neighbors of a vertex $i \in \Gamma$ is called its degree n_i . For functions v from the vertices of Γ to \mathbb{R} , we define the (normalized) Laplacian as

$$\Delta v(i) := v(i) - \frac{1}{n_i} \sum_{j, j \sim i} v(j). \quad (1)$$

(Note that this operator is different from, and in particular, has a different spectrum than the operator $Lv(i) := n_i v(i) - \sum_{j, j \sim i} v(j)$ usually studied in the graph theoretical literature as the (algebraic) graph Laplacian, see e.g. [9, 14, 21, 23, 8], but has the same spectrum as the Laplacian investigated in [10]. The normalized Laplacian is the operator underlying random walks on graphs, and it naturally incorporates a conservation law.)

We are interested in the spectrum of this operator as yielding important invariants of the underlying graph Γ and incorporating its qualitative properties. As in the case of the algebraic Laplacian, one can essentially recover the graph from its spectrum, up to isospectral graphs. The latter are known to exist, but are – arguably¹ – relatively rare and qualitatively quite similar in most respects

¹For example, most trees are not uniquely determined by their spectrum.

(see [33] for a survey). For a heuristic algorithm for recovering a graph from the spectrum of its algebraic Laplacian which can be easily modified for the normalized Laplacian, see [15].

We now recall some elementary properties, see e.g. [10, 20]. The normalized Laplacian, henceforth simply called the Laplacian, is symmetric for the product

$$(u, v) := \sum_{i \in V} n_i u(i) v(i) \quad (2)$$

for real valued functions u, v on the vertices of Γ . Δ is nonnegative in the sense that $(\Delta u, u) \geq 0$ for all u .

The eigenvalues of Δ therefore are real and nonnegative, the eigenvalue equation being

$$\Delta u - \lambda u = 0. \quad (3)$$

A nonzero solution u is called an eigenfunction for the eigenvalue λ . Since Γ has N vertices, the function space on which Δ operates is N -dimensional. Therefore, it has N eigenvalues; some of them might occur with multiplicity > 1 . The eigenfunctions for an eigenvalue λ constitute a vector space whose dimension equals the multiplicity of the eigenvalue λ . In the sequel, when we describe an eigenfunction, this is to be taken as some suitable element of this vector space. The smallest eigenvalue is $\lambda_0 = 0$, with a constant eigenfunction. This eigenvalue is simple because we assume that Γ is connected; in general, the multiplicity of the eigenvalue 0 equals the number of connected components, with the corresponding eigenfunctions being $\equiv 1$ on one and $\equiv 0$ on all other components. Returning to our case of a connected graph Γ , then

$$\lambda_k > 0 \quad (4)$$

for $k > 0$ where we order the eigenvalues as

$$\lambda_0 = 0 < \lambda_1 \leq \dots \leq \lambda_{N-1}.$$

After the brief discussion of the smallest eigenvalue, 0, we now turn to the largest one; here, we have

$$\lambda_{N-1} \leq 2, \quad (5)$$

with equality iff the graph is bipartite. Thus, a single eigenvalue determines the global property of bipartiteness. In fact, it is also true that a graph is bipartite iff whenever λ is an eigenvalue, then so is $2 - \lambda$. In other words, a bipartite graph has a spectrum that is symmetric about 1, and this characterizes bipartiteness. It is also instructive to look at the spectrum of particular graphs. For example, for a complete graph of N vertices, we have

$$\lambda_1 = \dots = \lambda_{N-1} = \frac{N}{N-1}, \quad (6)$$

that is, the eigenvalue $\frac{N}{N-1}$ occurs with multiplicity $N - 1$. Among all graphs with N vertices, this is the largest possible value for λ_1 and the smallest possible value for λ_{N-1} . Again, this spectral property fully characterizes complete

graphs.

The preceding examples concern exact values for the eigenvalues. In contrast, qualitative properties of a graph are usually characterized by inequalities for its eigenvalues, an issue that we shall return to below.

3 Eigenfunctions

When we think of a graph Γ representing biological data as a structure that has evolved from some simpler precursors, for example by joining smaller graphs into a larger one, or by duplicating certain sets of vertices in a precursor graph, it is important to find some indications of this process in the spectrum of Γ . It turns out that also certain properties of eigenfunctions can be useful here. We shall now describe some such aspects in formal terms (for some details, we refer to [6]).

In some cases, a solution u_k of the eigenvalue equation

$$\Delta u_k - \lambda_k u_k = 0 \quad (7)$$

can be localized, that is, be 0 outside a small set of vertices. In other cases, it has to be global, that is, be 0 only at relatively few vertices. These are qualitative notions, but they provide some insight into the behavior of graphs under certain operations as we shall now explore.

The considerations will depend on the eigenvalue equation (3), rewritten as

$$\frac{1}{n_i} \sum_{j \sim i} u(j) = (1 - \lambda) u(i) \text{ for all } i. \quad (8)$$

We observe that when the eigenfunction u vanishes at i , then also

$$\sum_{j \sim i} u(j) = 0. \quad (9)$$

The converse also holds, except for the case $\lambda = 1$ when (9) holds at all points regardless of whether the eigenfunction vanishes there or not.

We start with constructions that lead to localized eigenfunctions. We think of a motif as a small graph whereas the graph Γ is supposed to be large. This is not at all necessary for the subsequent constructions, but is in the spirit of the term “localized”.

1. **Motif joining:** Let Γ_0 be another graph, j_0 a vertex of Γ_0 , with an eigenvalue λ and an eigenfunction u^λ that vanishes at j_0 , i.e., $u^\lambda(j_0) = 0$. When we then form a graph $\bar{\Gamma}$ by identifying the vertex j_0 with an arbitrary vertex i of Γ , the new graph $\bar{\Gamma}$ also has the eigenvalue λ , with an eigenfunction that agrees with u^λ on Γ_0 and vanishes at the other vertices, that is, those coming from Γ . Thus, a motif Γ_0 can be joined to an existing graph with a preserved eigenvalue and a localized eigenfunction when the joining occurs at one (or several) vertices where that eigenfunction vanishes.

2. **Motif duplication:** Let Γ_1 be a motif in Γ , that is, a (small) subgraph of Γ , with vertices j_1, \dots, j_m . Let the function u on the vertex set of Γ_0 satisfy

$$\frac{1}{n_i} \sum_{j \in \Gamma_1, j \sim i} u(j) = (1 - \lambda)u(i) \text{ for all } i \in \Gamma_1 \text{ and some } \lambda. \quad (10)$$

Let $\bar{\Gamma}$ be obtained from Γ by doubling the motif Γ_1 , that is, by adding vertices i_1, \dots, i_m and their connections as in Γ_1 and connecting each i_α with all $i \notin \Gamma_1$ that are neighbors of j_α . Then the graph $\bar{\Gamma}$ possesses the eigenvalue λ with an eigenfunction u^λ that is localized on Γ_1 and its double; it agrees with u on Γ_1 , with $-u$ on the double of Γ_1 , and vanishes on the rest of $\bar{\Gamma}$. Thus, the eigenvalue λ is produced from motif duplication with symmetric eigenfunction balancing.

Not all eigenvalues possess localized eigenfunctions. Take cyclic graphs Γ_1, Γ_2 of lengths $4m - 1$ and $4n + 1$, for some positive integers m, n . Since the only cyclic graphs that admit the eigenvalue 1 are those of length $4k$, neither Γ_1 nor Γ_2 possesses the eigenvalue 1, but if we join them by identifying a vertex $i_0 \in \Gamma_1$ with a vertex $j_0 \in \Gamma_2$ the resulting graph Γ has 1 as an eigenvalue. An eigenfunction has the value 1 at the joined vertex, and the values ± 1 occurring always in neighboring pairs at the other vertices of Γ_1, Γ_2 , where the two neighbors of $i_0 = j_0$ in Γ_1 both get the value -1 , and the ones in Γ_2 the value 1. Since the multiplicity of the eigenvalue 1 on Γ is 1, there exists no other linearly independent eigenfunction for the eigenvalue 1. Thus, the local construction of joining two graphs at a single vertex here produces an eigenfunction that cannot be localized.

As another example, we can take any two graphs Γ_1, Γ_2 . Their disjoint union then has two components, and therefore, the multiplicity of the eigenvalue 0 is 2. One eigenfunction u_0 is $\equiv 1$ on Γ_1 and $\equiv 0$ on Γ_2 , and for the other one, v_0 , the roles of the components are reversed. When we now form a graph Γ by connecting some vertex $i_0 \in \Gamma_1$ to some vertex $j_0 \in \Gamma_2$ by an edge, then the multiplicity of the eigenvalue $\lambda_0 = 0$ becomes 1 because Γ is connected; the corresponding eigenfunction u is $\equiv 1$. However, when both Γ_1 and Γ_2 are large, the next² eigenvalue λ_1 of Γ is very small, and a corresponding eigenfunction is well approximated by one, v , that equals a positive constant on Γ_1 and a negative constant on Γ_2 (satisfying $\sum_{i \in \Gamma} n_i v(i) = 0$). Thus, u is a symmetric linear combination, v an antisymmetric one of the original eigenfunctions u_0, v_0 , and also the eigenvalues are close.

4 Properties of spectral plots and evolution hypotheses

Constructions like motif joining or duplication describe certain processes of graph formation that leave characteristic traces in the spectrum. This sug-

²assuming for simplicity that Γ_1, Γ_2 do not have small nontrivial eigenvalues themselves

gests that they can also serve useful roles for developing hypotheses about the evolution of a graph representing actual biological data. Of course, such hypotheses then need to be biologically plausible as well. Let us consider some examples. The simplest version of motif duplication is the doubling of a single vertex $j_1 \in \Gamma$. According to the general scheme, we add a new vertex i_0 and connect i_0 with all neighbors of j_0 . This generates an eigenvalue 1, with an eigenfunction u_1 that is localized at j_0 and i_0 , $u_1(j_0) = 1, u_1(i_0) = -1$. Thus, if the spectral plot of a graph has a high peak at the eigenvalue 1, a natural hypothesis is that this graph evolved via a sequence of vertex doubling. The next simplest case of a motif is an edge connecting two vertices j_1, j_2 . (10) then becomes

$$\frac{1}{n_{j_1}}u(j_2) = (1 - \lambda)u(j_1), \quad \frac{1}{n_{j_2}}u(j_1) = (1 - \lambda)u(j_2), \quad (11)$$

with the solutions

$$\lambda_{\pm} = 1 \pm \frac{1}{\sqrt{n_{j_1}n_{j_2}}}. \quad (12)$$

Thus, the duplication of an edge produces the eigenvalues λ_{\pm} . These are symmetric about 1. Also, when the degree of j_1 or j_2 is large, λ_{\pm} are close to 1. Thus, when the spectral peak at 1 is high, but not too sharp, and symmetric about 1, this is an indication that edge duplication has played some role in the evolution of the structure.

Next, we connect an edge between vertices j_1, j_2 to an existing graph Γ by connecting both j_1 and j_2 via an edge to some vertex $i_0 \in \Gamma$, or equivalently, we join a triangle with vertices j_0, j_1, j_2 to Γ by identifying j_0 with $i_0 \in \Gamma$. In that case, we produce the eigenvalue $3/2$. An eigenfunction u for the eigenvalue $3/2$ satisfies $u(j_1) = 1, u(j_2) = -1$, and vanishes elsewhere. Thus, again, it is localized. The same result obtains when we join the triangle by connecting j_0 and i_0 by an edge instead of identifying them. A high multiplicity of the eigenvalue $3/2$ may then generate the hypothesis that such processes of triangle joining repeatedly occurred in the evolution of the structure.

When in addition to the triangle j_0, j_1, j_2 another triangle k_0, k_1, k_2 is joined by identifying both j_0 and k_0 with $i_0 \in \Gamma$, we not only generate the eigenvalue $3/2$ with multiplicity 2, but also the eigenvalue $1/2$, with an eigenfunction $v(j_1) = v(j_2) = 1, v(k_1) = v(k_2) = -1$ and 0 elsewhere. Again, such a feature when prominently observed in a spectral plot may induce a corresponding hypothesis.

The described operations can also be of a global nature. For example, we can double the entire graph Γ ; when Γ consists of the vertices p_1, \dots, p_N , we take another copy Γ' with vertices q_1, \dots, q_N and the same connection pattern and connect each q_{α} also to all neighbors of p_{α} . The new graph $\bar{\Gamma}$ then has the same eigenvalues as Γ , plus the eigenvalue 1 with multiplicity N . This is biologically relevant, because there is some evidence for whole genome duplication [25, 29, 32]. However, protein-protein interaction networks do have a high multiplicity, but not of the order of half the system size [5]. This is readily explained

by subsequent mutations after the genome duplication that destroy the symmetry and thereby reduce the multiplicity of the eigenvalue 1. Also, since graph duplication does not change λ_1 and λ_{N-1} , the synchronization properties are not affected (see [20]).

5 Examples of spectral plots

We now exhibit spectral plots of different formal and biological networks. We convolve the eigenvalues with a Lorentz kernel, that is, we plot the graph of the function

$$f(x) = \sum_{\lambda_j} \frac{\gamma}{(\lambda_j - x)^2 + \gamma^2}$$

where the λ_j are the eigenvalues and we choose the parameter value $\gamma = .03$.³ In Figure 1, we see an Erdős-Rényi random network, a Strogatz-Watts small-world network, and a Barabási-Albert scale-free network. Each of these types has its very distinct shape, and this is not affected by varying the parameters underlying the construction schemes. In Figure 2, we then see a protein-protein interaction network and two neurobiological networks, and in Figure 3, we have a metabolic network, a food-web, and a transcription network. These are just examples, and choosing other examples from the same category yields very similar shapes. As we directly see, however, shapes of spectral plots for networks from different biological realms are very different from each other and from the formal networks, even though those have been suggested to capture important aspects of biological networks. Clearly, this indicates that for analysing biological networks, it does not suffice to rely on some generic formal construction scheme. Rather, one needs to analyse the specific aspects of specific biological realms through formal methods that are sufficiently rich to capture the essential qualitative features of that biological domain. In this paper, we have proposed spectral analysis as such a method.

References

- [1] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74, 2002, 47–97.
- [2] F.M. Atay, T. Bıyıkoglu, J. Jost, Synchronization of networks with prescribed degree distributions, IEEE Trans. Circuits and Systems I 53 (1), 2006, 92–98.

³In fact, we could as well take some other kernel here; the general formula is $f(x) = \sum_{\lambda_j} \int k(x, \lambda) \delta(\lambda - \lambda_j) d\lambda$ where $k(x, \lambda)$ is some kernel function. As an alternative to the Lorentz kernel, we could also take, e.g., a Gaussian kernel, or a piecewise constant kernel $k(x, \lambda) = \frac{1}{2\gamma}$ if $|x - \lambda| \leq \gamma$ and 0 else. The shape of the kernel is less important here than a careful choice of the parameter γ . For small γ , the plot obscures the global features, while for large γ , the details become too blurred.

- [3] F.M. Atay, T.Biyikoğlu, J.Jost, Network synchronization: Spectral versus statistical properties, *Phys.D*, to appear
- [4] F.M. Atay, J. Jost, A. Wende, Delays, connection topology, and synchronization of coupled chaotic maps, *Phys. Rev. Lett.* 92 (14), 2004, 144101.
- [5] A.Banerjee, J.Jost, Laplacian spectrum and protein-protein interaction networks, e-print: arXiv:0705.3373v1
- [6] A.Banerjee, J.Jost, On the spectrum of the normalized graph Laplacian, e-print: arXiv:0705.3772v1
- [7] A.-L. Barabási, R. A. Albert, Emergence of scaling in random networks, *Science* 286, 1999, 509–512.
- [8] T.Biyikoğlu, J.Leydold, P.Stadler, Laplacian eigenvectors of graphs, Springer LNM, to appear
- [9] B.Bolobás, *Modern graph theory*, Springer, 1998
- [10] F.Chung, *Spectral graph theory*, AMS, 1997
- [11] S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks*, Oxford, 2003.
- [12] G.Gladwell, E.Davies, J.Leydold, and P.Stadler, Discrete nodal domain theorems, *Lin.Alg.Appl.*336, 2001, 51-60
- [13] P. Erdős, A. Rényi, On random graphs, *Publicationes Mathematicae Debrecen*, 6, 1959, 290-297
- [14] C.Godsil, G.Royle, *Algebraic graph theory*, Springer, 2001
- [15] M. Ipsen, A. S. Mikhailov, Evolutionary reconstruction of networks, *Phys. Rev. E* 66(4), 2002
- [16] H. Jeong and B. Tombor and R. Albert and Z. N. Oltval and A. L. Barabási, The Large-Scale Organization of Metabolic Networks, *Nature*, 407(6804), 2000, 651-654
- [17] H. Jeong, S. P. Mason, A. L. Barabási, Z. N. Oltvai, Lethality and Centrality in Protein Networks, *Nature*, 411(6833), 2001, 41-42
- [18] J. Jost, *Mathematical methods in biology and neurobiology*, monograph, to appear
- [19] J. Jost, in: J.F.Feng, J.Jost, M.P.Qian (eds.), *Networks: from biology to theory*, Springer, 2007
- [20] J. Jost, M. P. Joy, Spectral properties and synchronization in coupled map lattices, *Phys.Rev.E* 65(1), 2002, 016201

- [21] R. Merris, Laplacian matrices of graphs – a survey, *Lin. Alg. Appl.* 198, 1994, 143-176
- [22] R. Milo and S. Shen-Orr and S. Itzkovitz and N. Kashtan and D. Chklovskii and U. Alon, Network Motifs: Simple Building Blocks of Complex Networks, *Science*, 298(5594), 2002, 824-827
- [23] B. Mohar, Some applications of Laplace eigenvalues of graphs, in: G.Hahn, G.Sabidussi (eds.), *Graph symmetry: Algebraic methods and applications*, pp. 227-277, Springer, 1997
- [24] M. Newman, The structure and function of complex networks, *SIAM Review* 45, 2003, 167–256
- [25] S.Ohno, *Evolution by Gene Duplication*, Springer, 1970
- [26] L.M. Pecora, T.L. Carroll, Synchronization in chaotic systems, *Phys. Rev. Lett.* 64, 1990, 821–824
- [27] A.Pikovsky, M.Rosenblum, J.Kurths, *Synchronization – A Universal Concept in Nonlinear Science*, Cambridge University Press, Cambridge, 2001
- [28] S. S. Shen-Orr and R. Milo and S. Mangan and U. Alon, Network Motifs in the Transcriptional Regulation Network of Escherichia Coli, *Nature Genetics*, 31(1), 2002, 64-68
- [29] A. Wagner, Evolution of Gene Networks by Gene Duplications - a Mathematical-Model and Its Implications On Genome Organization, *Proc. Nat. Acad. Sciences USA* 91(10), 1994, 4387-4391
- [30] D. J. Watts, S. H. Strogatz, Collective Dynamics of 'Small-World' Networks, *Nature*, 393(6684), 1998, 440-442
- [31] J. G. White and E. Southgate and J. N. Thomson and S. Brenner, The Structure of the Nervous-System of the Nematode *Caenorhabditis-Elegans*, *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 314(1165), 1986, 1-340
- [32] K. H. Wolfe, D. C. Shields, Molecular Evidence for an Ancient Duplication of the Entire Yeast Genome, *Nature*, 387(6634), 1997, 708-713
- [33] P.Zhu, R.Wilson, A study of graph spectra for comparing graphs, <http://cms.brookes.ac.uk/staff/PhilipTorr/BMVC2005/papers/paper-57-162.html>

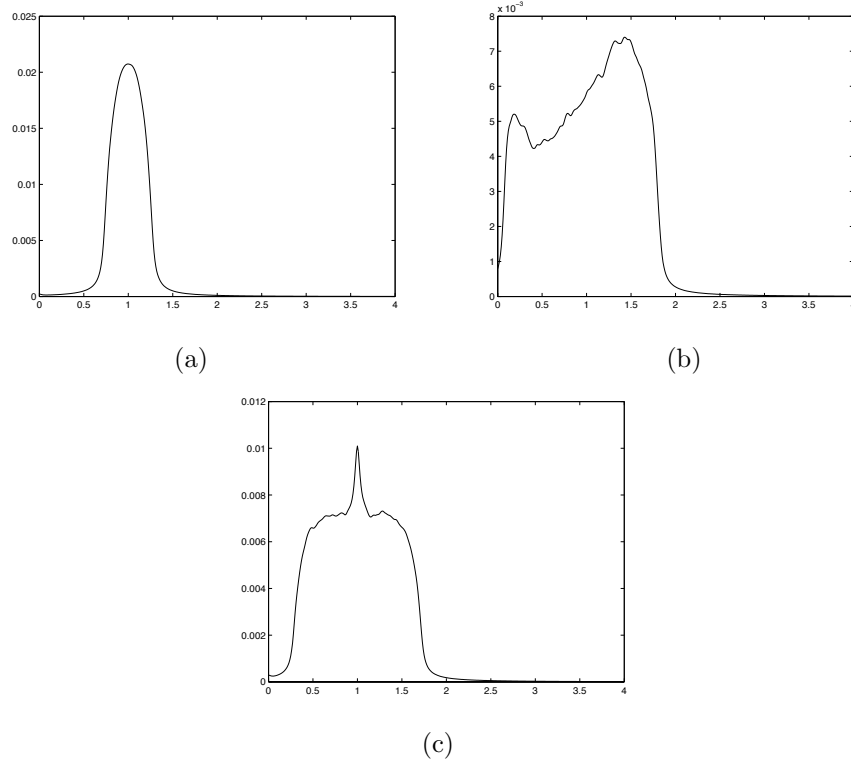


Figure 1: Spectral plots of generic networks. (a) Random network by Erdős and Rényi's model [13] with $p = 0.05$. (b) Small-world network by Watts and Stogatz model [30] (rewiring a regular ring lattice of average degree 4 with rewiring probability 0.3). (c) Scale-free network by Albert and Barabási model [7] ($m_0 = 5$ and $m = 3$). Size of all networks is 1000. All figures are plotted with 100 realization.

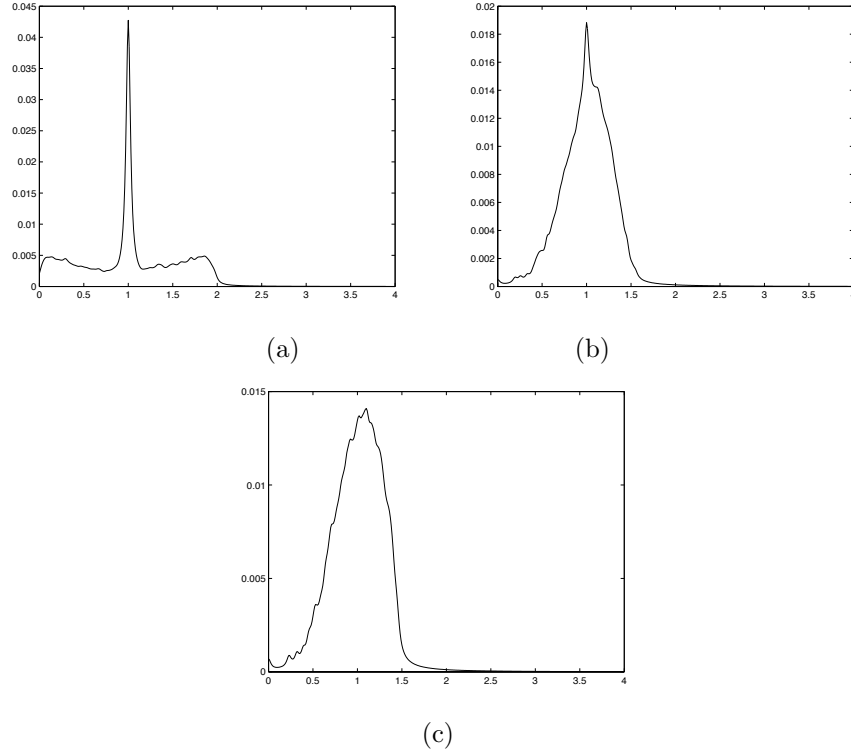


Figure 2: Spectral plots of (a) protein-protein interaction network of *Saccharomyces cerevisiae* (yeast). Network size is 1458. Data downloaded from <http://www.nd.edu/~networks/> and data used in [17] [download date: 17th September, 2004] (b) neuronal connectivity of *C. elegans*. Size of the network is 297. Data used in [30, 31]. Data Source: <http://cdg.columbia.edu/cdg/datasets/> [Download date: 18th Dec. 2006]. (c) neuronal connectivity of *C. elegans* from the animal JSH, L4 male in the nerve ring and RVG regions. Network size is 190. Data source: Data is assembled by J. G. White, E. Southgate, J. N. Thomson, S. Brenner [31] and was later revisited by R. M. Durbin (Ref. <http://elegans.swmed.edu/parts/>). [Download date: 27th Sep. 2005].

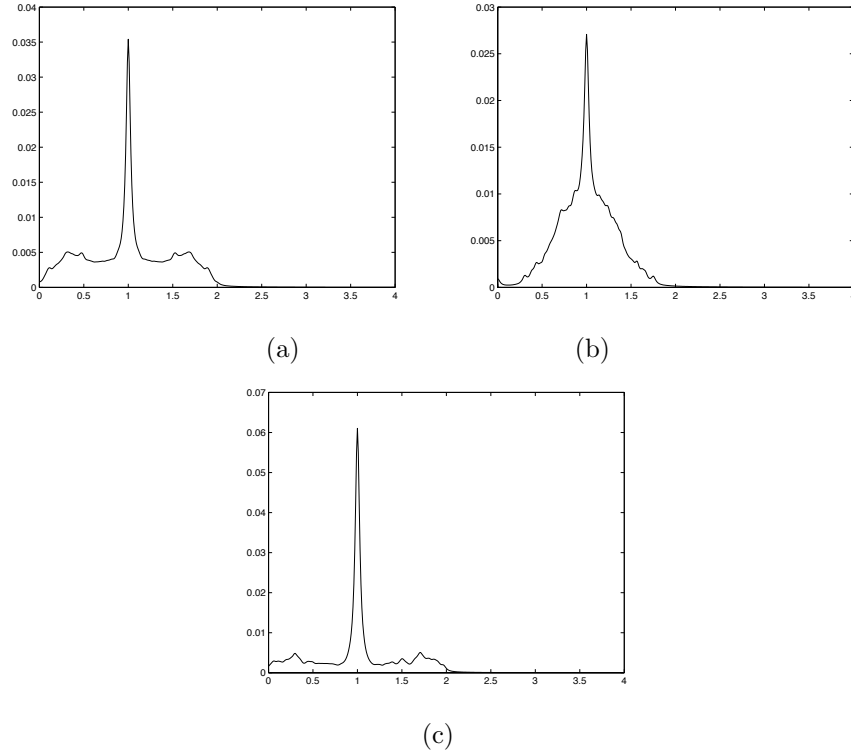


Figure 3: Spectral plots of (a) metabolic network of *A. pernix*. Network size is 490. Here nodes are substrates, enzymes and intermediate complexes. Data used in [16]. Data Source: <http://www.nd.edu/~networks/resources.htm/>. [Download date: 22nd Nov. 2004]. (b) food-web from "Ythan estuary". Size of the network is 135. Data downloaded from <http://www.cosin.org/>. [Download Date 21st December, 2006]. (c) transcription network of *E. coli*. Size of the network is 328. Data source: Data published by Uri Alon (<http://www.weizmann.ac.il/mcb/UriAlon/>). [Download date: 13th Oct. 2004]. Data used in [22, 28].